Inferences for a Mean Vector

* Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ are a random sample from a $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

* Generally both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown.

* In this section we would like to make inferences about $\boldsymbol{\mu}$ so $\boldsymbol{\Sigma}$ is a nuisance parameter.

* Among the inferences we would like to make are
  - Test $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ for a known vector $\boldsymbol{\mu}_0$.
  - Construct a confidence region for $\boldsymbol{\mu}$.

Recap of Univariate Inference

* Suppose that $p = 1$ and we wish to make inference about μ.

* Inference is generally based on

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

* W.S. Gosset (Student) showed that T has a t distribution with $n - 1$ degrees of freedom.

* This result allows us to test hypotheses about μ and construct confidence intervals.

Testing Hypotheses

* Suppose that we wish to test $H_0 : \mu = \mu_0$ then, when this hypothesis is true,

$$T_0 \;=\; \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \;\sim\; t_{n-1}$$

* If we let t_0 be the observed value of T_0 in our sample then a p-value of the test is

$$p \;=\; \mathsf{P}(|T_0| > |t_0|) \;=\; \mathsf{P}\left(T_0^2 > t_0^2\right)$$

* We reject H_0 if p is small.

Confidence Intervals

* Suppose that we decide to reject H_0 if $p < \alpha$.

* This is equivalent to saying we reject if

$$|t_0| \;=\; \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| \;<\; t_{n-1}(\alpha/2)$$

  where $t_{n-1}(\alpha/2)$ is the $1 - \alpha/2$ quantile of the t_{n-1} distribution.

* If we do NOT reject H_0 then we are saying that μ_0 is a plausible value of μ.

* There will be many plausible values!

Confidence Intervals

* The collection of all plausible values is

$$\left\{\mu_0 \in \mathbb{R} : \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| \ < \ t_{n-1}(\alpha/2)\right\}$$

* An alternative way of writing this is

$$\left\{\mu_0 \in \mathbb{R} : \overline{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \ < \ \mu_0 \ < \ \overline{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right\}$$

* The interval

$$\left(\overline{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}, \quad \overline{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right)$$

is a $100(1-\alpha)\%$ Confidence Interval For μ.

t and F Distributions

Theorem 17

Suppose that $Z \sim N(0,1)$, $X \sim \chi^2_k$ and Z and X are independent. Then

$$T = \frac{Z}{\sqrt{X/k}} \sim t_k$$

Theorem 18

Suppose that $X_1 \sim \chi^2_{k_1}$, $X_2 \sim \chi^2_{k_2}$ and X_1 and X_2 are independent. Then

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F_{k_1,k_2}$$

Theorem 19

If $T \sim t_k$ then $F = T^2 \sim F_{1,k}$.

Extending to the Multivariate Setting

* An equivalent to the T statistic for two-sided inference is the statistic

$$T^2 \;=\; \frac{n(\overline{X}-\mu)^2}{S^2} \;\sim\; F_{1,n-1}$$

* We can write this as

$$T^2 \;=\; n(\overline{X}-\mu)S^{-1}(\overline{X}-\mu)$$

* An obvious generalization of this to multivariate setting is then

$$T^2 \;=\; n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})^t\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})$$

* This statistic is known as the Hotelling's T^2 Statistic.

Distribution of T^2

* The T^2 statistic given on the previous page is named named after Harold Hotelling, an American mathematical statistician.

* In 1931, Hotelling examined this extension of the Student's t statistic and showed that the sampling distribution of T^2 is proportional to an F distribution.

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

* Note that when $p = 1$ this reduces to the result given earlier.

Application To Multivariate Testing

* Suppose that we wish to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ for some specified $\boldsymbol{\mu}_0$.

* Define the test statistic

$$T_0^2 \;=\; n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \boldsymbol{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

  and let t_0^2 be the observed value.

* Then we can test H_0 by calculating the p-value

$$p \;=\; \mathsf{P}\left(T_0^2 > t_0^2\right)$$

* Rejecting for $p < \alpha$ is equivalent to rejecting for

$$t_0^2 \;>\; \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

  where $F_{p,n-p}(\alpha)$ is the $1-\alpha$ quantile of the $F_{p,n-p}$ distribution.

Properties of the Test

* If we reject H_0 then this means that At least one of the components of μ is not equal to the corresponding component of μ_0.

* The validity of the test does rely on the multivariate normality assumption so this should always be checked before applying the test.

* An interesting result is that if we let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ for a non-singular $p \times p$ matrix $\boldsymbol{A}$ and constant vector $\boldsymbol{b} \in \mathbb{R}^p$ then the statistic for testing $\boldsymbol{\mu}_{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{\mu}_0 + \boldsymbol{b}$ based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ is exactly the same as that for testing $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ based on the original sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

Likelihood Ratio Tests

* A general way of testing composite hypotheses.

* Suppose that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the parameter of the distribution and the likelihood based on a random sample is

$$L(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\boldsymbol{X}}(\boldsymbol{x}_i; \boldsymbol{\theta})$$

* Then the likelihood ratio test statistic of $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ is

$$\Lambda = \frac{\max\limits_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}{\max\limits_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}$$

* The likelihood ratio test procedure is then to reject H_0 if $\boldsymbol{\Lambda} < c$ where c is chosen so that

$$\mathsf{P}(\boldsymbol{\Lambda} < c; \boldsymbol{\theta} \in \boldsymbol{\Theta}_0) = \alpha$$

Application to the Normal Mean

* In this case we have $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and we know from earlier work that

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) \;=\; L(\overline{\boldsymbol{x}},\widehat{\boldsymbol{\Sigma}}) \;=\; \frac{1}{(2\pi)^{np/2}\left|\widehat{\boldsymbol{\Sigma}}\right|^{n/2}}\mathrm{e}^{-np/2}$$

* Similarly we can see that

$$\begin{aligned}\max_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}) \\ &= L(\boldsymbol{\mu}_0,\widehat{\boldsymbol{\Sigma}}_0) \\ &= \frac{1}{(2\pi)^{np/2}\left|\widehat{\boldsymbol{\Sigma}}_0\right|^{n/2}}\mathrm{e}^{-np/2}\end{aligned}$$

where

$$\widehat{\boldsymbol{\Sigma}}_0 \;=\; \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^t$$

Application to the Normal Mean

* Using these results we see that the likelihood ratio statistic can be written as

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{n/2} = \Lambda_1^{n/2}$$

* The quantity $\Lambda_1 = |\hat{\Sigma}|/|\hat{\Sigma}_0|$ is known as Wilk's Lambda.

* Obviously rejecting when Λ is small is equivalent to rejecting when Λ_1 is small so it suffices to consider Λ_1 as the test statistic.

Relationship with Hotelling's Test

Theorem 20

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let

$$T_0^2 \;=\; n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \boldsymbol{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

be the Hotelling's T^2 statistic for testing $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and let

$$\Lambda_1 \;=\; \frac{\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_0|} \;=\; \frac{\left|\sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t\right|}{\left|\sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^t\right|}$$

Then

$$\Lambda_1 \;=\; \left(1 + \frac{T^2}{n-1}\right)^{-1}.$$

Relationship with Hotelling's Test

* The previous theorem implies that rejecting for small values of Λ_1 is exactly equivalent to rejecting for large values of T_0^2 and so the two tests are equivalent.

* It also shows that we do not need to invert S to get T^2 since we can simply use the determinants

$$T_0^2 = (n-1)\left(\frac{\left|\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^t\right|}{\left|\sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^t\right|}-1\right)$$

Confidence Ellipsoids for μ

- Based on the Hotelling's T^2 we can find a confidence region for $\boldsymbol{\mu}$.

- We do not reject $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ at a significance level α if

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^t \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \; < \; \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

- Hence we can use the confidence ellipsoid

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^t \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \; < \; \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

as a set of plausible values of μ with confidence level $100(1-\alpha)\%$.

Axes of the Confidence Ellipsoid

* From the work we did earlier we can see that the axes of the confidence ellipsoid are parallel to the eigenvectors of S.

* Furthermore the length of the j^{th} axis is proportional to the j^{th}-largest eigenvalue λ_j.

* The j^{th} principal axis is then

$$\overline{\boldsymbol{x}} \pm \sqrt{\frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)}\; \boldsymbol{e}_j$$

Inference for a Linear Combination

* We are often interested in making inference about $\boldsymbol{a}^t\boldsymbol{\mu}$ for some vector $\boldsymbol{a} \in \mathbb{R}^p$.

* When we are truly only interested in a single linear combination $\boldsymbol{a}^t\boldsymbol{\mu}$ we can use the univariate statistic

$$T = \frac{\sqrt{n}(\boldsymbol{a}^t\overline{\boldsymbol{x}} - \boldsymbol{a}^t\boldsymbol{\mu})}{\sqrt{\boldsymbol{a}^t\boldsymbol{S}\boldsymbol{a}}} \sim t_{n-1}$$

* Hence we get the $100(1-\alpha)\%$ confidence interval

$$\boldsymbol{a}^t\overline{\boldsymbol{x}} \pm t_{n-1}(\alpha/2)\sqrt{\frac{\boldsymbol{a}^t\boldsymbol{S}\boldsymbol{a}}{n}}$$

* Usually, however, we are interested in many different linear combinations.

Simultaneous Inference

* Suppose that we may be interested in any linear combination.

* Individually intervals can be based on

$$T^2(\boldsymbol{a}) = \frac{n(\boldsymbol{a}^t\overline{\boldsymbol{x}} - \boldsymbol{a}^t\boldsymbol{\mu})^2}{\boldsymbol{a}^t\boldsymbol{S}\boldsymbol{a}} \leqslant c^2$$

where c^2 is chosen such that

$$\mathsf{P}\left(T^2(\boldsymbol{a}) \leqslant c^2 \text{ for any } \boldsymbol{a}\right) = \mathsf{P}\left(\max_{\boldsymbol{a}\in\mathbb{R}^p} T^2(\boldsymbol{a})\right) = 1 - \alpha$$

* It is possible to show that the Hotelling's T^2 maximizes $T^2(\boldsymbol{a})$ over all $\boldsymbol{a} \in \mathbb{R}^p$.

* Hence we get the simultaneous intervals

$$\boldsymbol{a}^t\overline{\boldsymbol{x}} \pm \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)\boldsymbol{a}^t\boldsymbol{S}\boldsymbol{a}}$$

Some Useful Linear Combinations

* For an individual component, $\boldsymbol{\mu}_j$ of $\boldsymbol{\mu}$ we get

$$\overline{x}_j \ \pm \ \sqrt{\frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) s_{jj}}$$

* For $\mu_j - \mu_k$ we get the interval

$$\overline{x}_j - \overline{x}_k \ \pm \ \sqrt{\frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) \left(s_{jj} + s_{kk} - 2s_{jk}\right)}$$

The Bonferroni Method

* In many situations we are not interested in all possible linear combinations.

* We may only be interested in a small subset of linear combinations of a specific type.

* For example maybe we are only interested in intervals for the p components of $\boldsymbol{\mu}$.

* In such cases, the simultaneous intervals given are too wide since they are designed to give the correct coverage over all possible $\boldsymbol{a}$.

The Bonferroni Method

* Suppose that we are only interested in m (pre-specified) linear combinations.

* To get an overall coverage of at least $1-\alpha$ for these intervals we can use the univariate intervals but adjust the critical value used.

* The Bonferroni Correction for Multiple Inference says that instead of using $\alpha/2$ for each univariate interval we should use $\alpha_k/2$ where $\sum \alpha_k = \alpha$

* Typically we take $\alpha_k = \alpha/m$ for $k = 1, \ldots, m$.

* Hence we get the intervals

$$\boldsymbol{a}_k^t \overline{\boldsymbol{x}} \; \pm \; t_{n-1}(\alpha/(2m))\sqrt{\frac{\boldsymbol{a}_k^t \boldsymbol{S} \boldsymbol{a}_k}{n}}$$

Large Sample Inferences

* Large sample testing can be based on the result that

$$T_0^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0) \xrightarrow{d} \chi_p^2$$

when $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

* From this we can get the confidence ellipsoid

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^t \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) < \chi_p^2(\alpha)$$

* Simultaneous confidence intervals can be similarly defined.

* Univariate results (including Bonferroni corrected ones) come from replacing t quantiles with standard normal quantiles.

* These results will hold provided n and $n - p$ are large.